Business Analytics Practicum - Groups 4 and 9

CHILD POVERTY ACTION LAB

An Analysis of Family Sub-indices and their impact on Child Poverty

Prepared by,
Bandarupalli, Sruthi
Bijani, Harshul
Cai, Chaoyu
Enugala, Amitesh
Li, Shiyou
Mittukolu, Raagni
Nidamanuru, Vinutan
Shaik, Ramiz Raja
Sharma, Ritu
Taneja, Vishal
Vaidyalingasarma, Gautam
Varma, Atul Kumar
Zhen, Kuo

Under the guidance of Professor Ron Bose.

ABOUT CHILD POVERTY ACTION LAB

Child poverty is a moral issue. Child Poverty is an economic issue. Child poverty is a Dallas issue. Children living in poverty are 3.5 times more likely to experience repeated trauma than their affluent peers. Due to the complexity of child poverty within the context of siloed resources, historical disparity, and vested interests, coordination across stakeholders is more difficult and necessary than ever. CPAL is built on the premise that the moral and economic imperative to reduce child poverty can incite new levels of collaboration and maximize the impact potential of our shared resources through the following theory of change: The right mix of systems and resources, must reach the right families and neighborhoods to mitigate the lasting effects of poverty and break the cycle for the next generation. CPAL's role is to build a shared understanding of the problem and the ecosystem, design solutions for scale, identify winnable milestones for measurable impact, rally demand and co-ordination for highest potential solutions, and embrace course corrections.

PROJECT SCOPE & OBJECTIVE

This project aims at identifying indicators that would represent the level to which family plays a role in child poverty in a given census tract. We sim at creating a score for all census tracts in Dallas. This score would tell us how well the census tract is doing in terms of family.

TECHNICAL DETAILS

SOFTWARES USED

For this project, we have used Python for data exploration, data cleansing, merging, regression, and index calculation. Tableau for data visualization.

DATA SOURCES

Census Tract, data.texas.gov, diversitydatakids.org

For each indicator, we collected data of four counties: Dallas, Collin, Denton, Tarrant.

INDICATORS USED

Following are the indicators we used to calculate the Family score for all census tracts. We have used a total of 11 indicators to arrive at the following. A brief description of the indicators is given below.

01 **ORPHAN CHILD**

Count of children w/o biological parents (ACS)

02 **AVG_FAMILY_SIZE**

Average family size (ACS)

03 **POORCHILD**

Number of children in poverty driven families (ACS)

04 **N_OF_OWNERS**

Number of Houseowner (ACS)

05 **DISABILITY**

Population under 18 years with a self-care difficulty (ACS)

06 **BOTH_PARENTS**

Number of two parents household (ACS)

07 **SINGLE_MOTHER**

Number of single mother household (ACS)

08 **SINGLE_FATHER**

Number of single father household (ACS)

09 **EDUC**

Number of people who are educated (ACS)

10 **INCOME**

Average income level (ACS)

11 **POVERTY**

Number of families living in poverty (ACS)

For 2018, the federal poverty level for individuals starts at $12140,increased by $4320 for each additional person

REASONS FOR CHOOSING THE INDICATORS

The following segment gives a clear reasoning behind choosing the indicators for computing the family index.

1. ORPHAN CHILD & POOR CHILD

The status of adults in a family can influence the future status of children, both directly (e.g., adults creating social and economic networks that connects youth to better lifestyle opportunities), and indirectly (e.g., adult's conditions and attitudes can shape or reinforce child's aspirations and decision-making in future).

We think that if a child is orphaned or if a child doesn't get proper care during early childhood days, this initial situation of poverty will make the child struggle for basic necessities in life. Due to this reason, we collected data on orphans and poor children as these two indicators could help us identify child poverty.

2. AVERAGE FAMILY SIZE

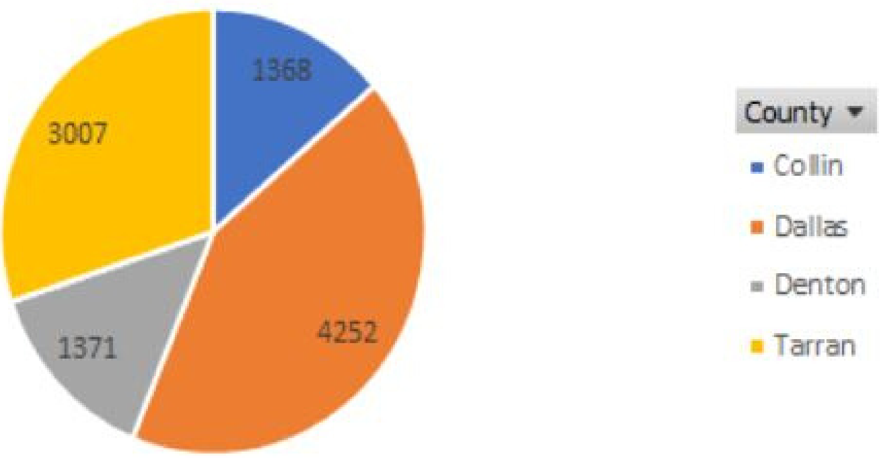
When the size of the family grows, so does the difficulty in supporting the family in most cases. A child who is part of a family with 2 adults and 2 children is often at an advantage compared to a child who is part of a family with 2 adults and 6 children. Lesser the number of children, better the attention they get from parents. For this reason, we thought average family size could be a very good indicator in computing the family index of all the census tracts. It could give us insights about families that are living a quality life and families that are potentially living a low quality life.

3. NUMBER OF HOUSE OWNERS

- The status of house ownership in a family can influence the future status of children.
- This indicator aims to measure the number of houses in a census tract that are owned versus the number of houses in that census tract that are rented.
- We believe that if the parents of the kids own houses, they are in a better position to provide a stable life to their kids in terms of access to resources, education, lifestyle, etc.

4. DISABILITY

- According to the U.S Census Bureau, there were nearly 40 million american with disabilities in the US. Disabled Americans earn less than those without a disability. Those with a disability earned a median of $21,572 in 2015, less than 70% of the median earnings for those without a disability ($31,872).
- According to the data collected by American Community Survey, there are 9998 kids (under 18 years old) with self- care difficulties in the Dallas area. Dallas County has the largest number, and Collin county has least.



5. FAMILY TYPE

Family of origin refers to the significant caretakers and siblings that a child grows up with, or the first social group a child belongs to, which is often a child's biological family or an adoptive family. Our early experiences have a major influence on how we see ourselves, others and the world and how we cope and function in our daily living.

- Based on the above conclusion, we collected the data about the parents marriage status including both_parent, single mother and single father.
- Since these three indicators are measuring the goodness of one single family environment, we decide to combine these three into one indicator called FamilyType to perform dimension reduction.

6. EDUCATION

- The number of adults aged 25 years and older who have completed a Bachelor's degree or higher, for the total population and by race/ethnicity.
- This indicator is mainly used for calculating the type score for the family.

7. HOUSEHOLD INCOME

Household income is indicative of both access to and perceptions of opportunity, and plays a powerful role in shaping individual choices and opportunities for the advancement of children. Neighborhoods high in economic resources have more financial resources to invest into amenities that depend on local funding, such as schools, parks, and after-school programs.

We think that based on the household income, a child can be provided with access to quality development. If a child doesn't belong to a household with sufficient income, decisions of life are compromised which could result in low quality living for the child.

METHODOLOGY

1. Data Standardization:

In order to scale our data to the range of 0 to 1, we decide to utilize Minmax scaling method for data preprocessing. And once we have the final index value, we still use this method and multiply 100 to scaling final index to the range of 0-100.

2. Run a collinearity test on the indicators:

For the collinearity check, we directly use correlation matrix to detect the collinearity relationship during these indicators.

3. Weight Calculation Method

i) <u>The Pearson Correlation:</u>

During the collinearity test, we observed that both parent, single mother, single father these three sub-indicators have relatively high correlation coefficient to each other. Therefore, we decide to combine these three sub-indicators together to construct Family Type indicator.

a. View both parent, single mother, single father as independent variables

b. Run the Pearson Correlation for independent variables against income, education and poverty separately

c. Now we have 3 numbers for every different dependent variable, income, education, poverty

d. Take average of the sub-indicators corresponding to the number. For example, if we have both parent equal to 0.03, 0.46, 0.40 in these three times, we will take (0.03+0.46+0.40)/3 = 0.3 as the result

e. Now we will get 3 values 1 each for single moth, single father and both Total is 1.18. So now do both/1.18, single moth/1.18 and single father/1.18. This is the final weight and construct the Family Type based on these weightages

ii) <u>Linear Regression Model Coefficient:</u>

a. Now regress income against type score and other 5 indicators. Adjust- R2 is 0.42

b. Family Type has highest weight

c. Sum up the coefficient of each indicators to get the total weight

d. Assigning weights to each indicators base on their coefficient

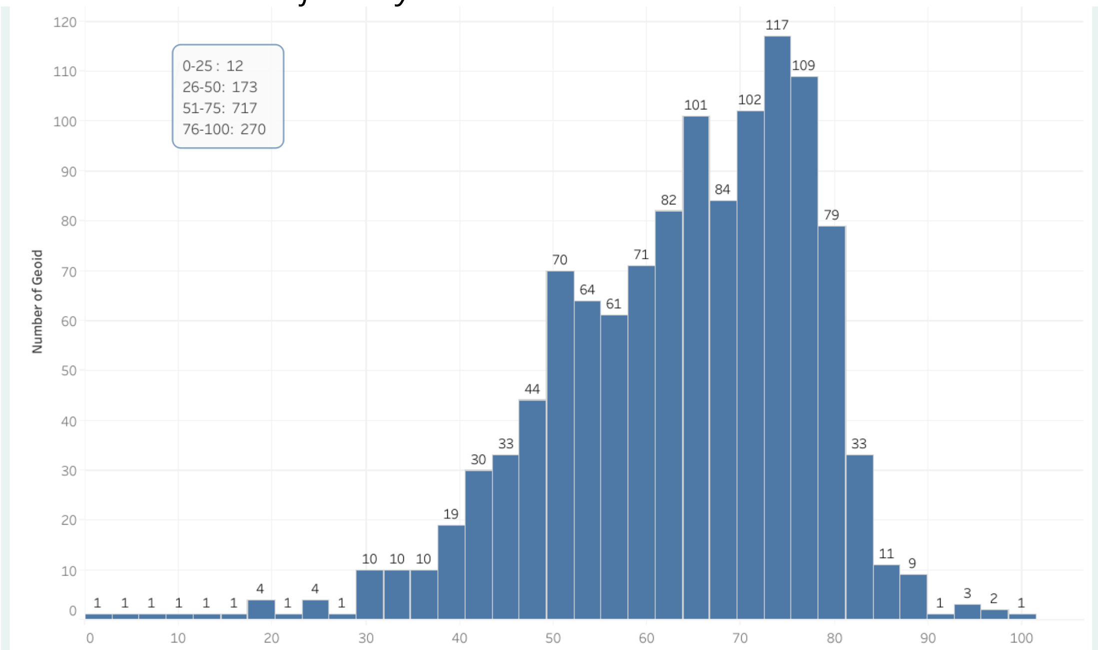
e. Combining indicators base on the weights to get the final score.

f. Scaling the final score to 0-100 scale.

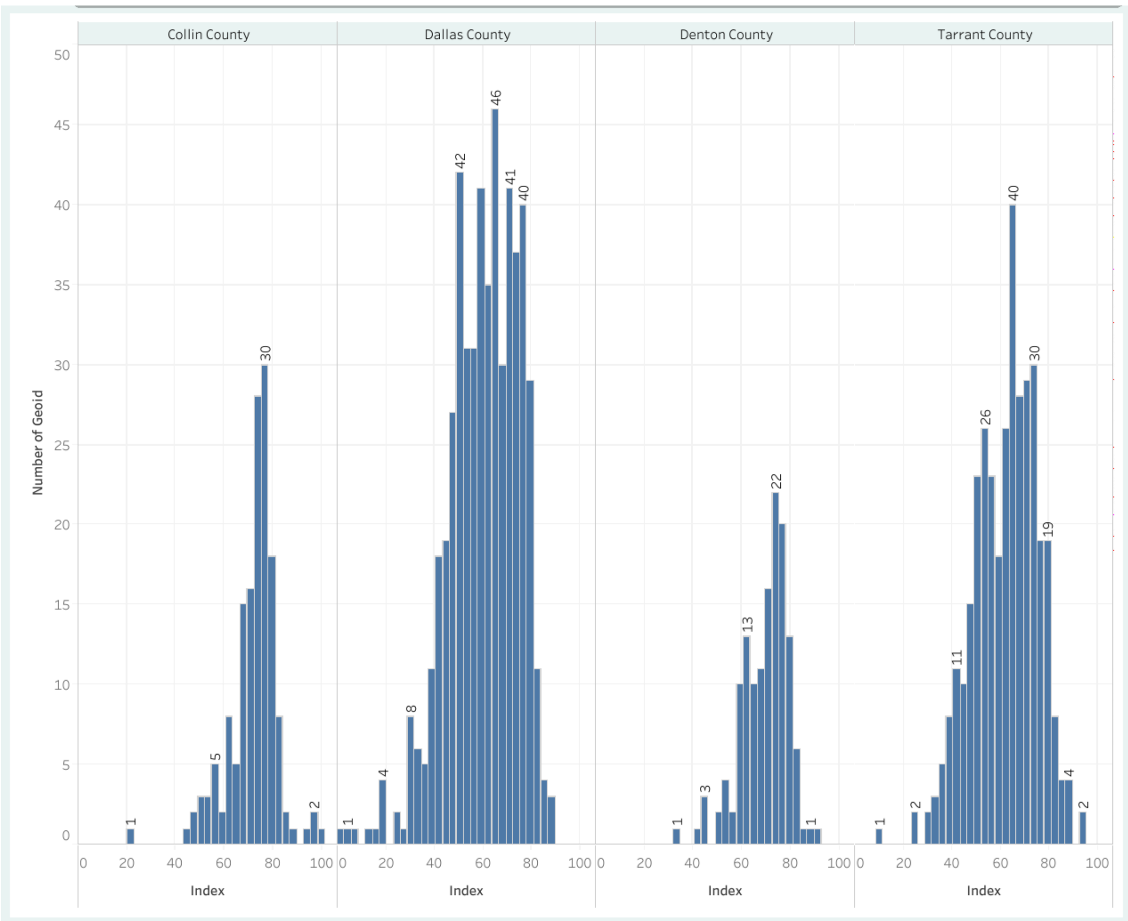
We have provided an appendix of all the formulae that we used for the calculations

RESULTS

Overall Distribution of Family Scores:



County wise distribution of Family Scores:



APPENDIX

Given below are the list of formulae we have used in our methodology.

a) Typescore:

Combining both, single father and single mother together with pearson correlation weight, we named this new indicator Typescore

```
df_type = pd.concat([X, family_scaled['Poverty']], axis=1, sort=False)
df_type.corr()
```

```
df_ty3 = pd.concat([X, family_scaled['Edu']], axis = 1)
df_ty3.corr()
```

```
df_ty4 = pd.concat([X, family_scaled['Income']], axis = 1)
df_ty4.corr()
```

TypeScore = 0.33*family_scaled['Both_Parents'] - 0.46*family_scaled['Single_Mother'] - 0.21*family_scaled['Single_Father']

b) Standardization

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaled_df = scaler.fit_transform(df)
scaled_df = pd.DataFrame(scaled_df, columns=df.columns)
```

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
f = np.array(family_scaled['final']).reshape(-1,1)
final_score = scaler.fit_transform(f)
```

c) Final Family Score

family_scaled['final'] = 0.10*family_scaled['Orphanchild'] - 0.14*family_scaled['Avg_Family_size'] - 0.12*family_scaled['PoorChild'] - 0.01*family_scaled['Disablity'] + 0.05*family_scaled['N_of_Owners'] + 0.57*family_scaled['TypeScore']